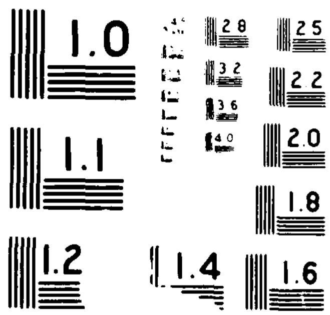
AD-A192 359 WORKING MEMORY CAPACITY: AN INDIVIDUAL DIFFERENCES APPROACH(U) SOUTH CAROLINA UNIV COLUMBIA DEPT OF PSYCHOLOGY R W ENGLE 11 FEB 88 AFOSR-TR-88-0265 1/1

UNCLASSIFIED AFOSR-87-0069 F/G 5/8 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.2
1.4
1.6

(2)

PSYCHOLOGY

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 88-0265
6a NAME OF PERFORMING ORGANIZATION Randall W. Engle Department of Psychology University of South Carolina / 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research/NL
6c ADDRESS (City, State, and ZIP Code) Columbia, SC 29208	7b ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, DC 20332-6448
8a NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR / 8b OFFICE SYMBOL (If applicable) NL	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR #870069
8c ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, DC 20332-6448	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F / PROJECT NO. 2313 / TASK NO. A4 / WORK UNIT ACCESSION NO.

11 TITLE (Include Security Classification)
Working Memory Capacity: An Individual Differences Approach

12 PERSONAL AUTHOR(S) Randall W. Engle

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
First Annual Technical	FROM Jan 1,87 TO Jan 1,88	February 11, 1988	64

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Working memory, capacity, individual differences, reading comprehension.

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The literature on individual differences in working memory capacity and its importance to reading comprehension is reviewed. The current project studies the relationship between the way working memory capacity is measured and comprehension. Five experiments are reported demonstrating that a significant relationship between the WM measure and comprehension does not depend on the background component of the WM measure being reading-related. While the predictability of the WM measure was higher if words were the items being recalled in the span task, it made no difference whether the background component was reading sentences or performing arithmetic operations. Even the simple word span significantly predicted comprehension if serial rather than free recall was required on the span component. The complex sentence word span appears to be influenced by the length of the words being recalled in the same way as has been repeatedly demonstrated for the simple span.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Alfred R. Fregly	22b TELEPHONE (Include Area Code) (202) 767-5021 / 22c OFFICE SYMBOL NL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

1) FEB 1988

Summary

Five experiments are described that study the relationship between measures of working memory and reading comprehension. Two experiments investigated whether the complex span measure must be similar to the reading comprehension task to be predictive of comprehension. The correlations found between reading comprehension and two reading-related complex spans was similar to those found between two arithmetic-related complex spans and comprehension. The relationship remained significant when quantitative skills were factored out. The simple digit and word spans (measured without a background task) did NOT correlate with reading comprehension. The complex span/comprehension correlations were a function of the difficulty of the background task. When the difficulty level of the reading-related or arithmetic-related background tasks was moderate, the span/comprehension correlations were higher in magnitude than when the background tasks were simple or very difficult.

The third experiment showed that if serial recall was required in the span tasks, simple word span did significantly predict reading comprehension but not as well as the sentence span. The fourth experiment showed that the ordering of list lengths in the span tasks had little influence on the correlation between span scores and comprehension. The fifth experiment is the first in a series investigating variables whether variables that influence simple word span also influence the sentence word span. This study demonstrated that the word length has the same effect on the sentence span task as on the simple word span.

Table of Contents

Working memory capacity: An individual differences approach

A hallmark of models of information processing since the early work of Shannon & Weaver (1949) has been the focus on capacity limitation. Most models attribute this limitation to some combination of short-term memory and attention (Broadbent, 1958). Presumably, the limited capacity of short-term memory affects our performance in cognitive tasks like reading or listening comprehension and general problem solving (Kintsch & van Dijk, 1978; Newell & Simon, 1972).

The digit span has been assumed to reflect output from short-term memory and is a ubiquitous component of intelligence tests (Wechsler, 1944). However, it does not correlate well with performance on such higher level tasks as reading comprehension (Perfetti & Lesgold, 1977) or even the amount of information estimated to be represented in primary or secondary memory (Martin, 1978).

Early theories, e.g., Waugh & Norman (1965), viewed short-term memory as a fixed number of slots or bins. Baddeley & Hitch (1974) argued that this focused too much on the storage functions of short-term memory and not enough on the processing functions. They preferred the name working memory and argued for the importance of both storage and processing in a functional analysis of the working memory system. Baddeley & Hitch proposed three structural components through which information is processed: (1) a central executive, (2) an articulatory loop, and (3) a visuo-spatial scratch pad. The articulatory loop and the visuo-spatial scratch pad are maintenance systems controlled by the central executive, which is a flexible work space with limited capacity. Part of this limited capacity is used for processing incoming information with the remainder used for storage of the products resulting from that processing.

Although other models of WM have been developed (e.g., Klapp, Marshburn & Lester, 1983; Brainerd & Kingma, 1984, 1985; Case, 1974; Kintsch & van Dijk, 1978), they all assume a limitation in the amount of information that can be kept active at any given time. Further, it has been generally assumed that this limitation affects consequent processing, i.e., that higher level processing is limited to some extent by the limitations of WM.

Daneman & Carpenter (1980, 1983) demonstrated the importance of the WM limitation in reading comprehension. They hypothesized that WM is used to represent the strategies and skills used in reading with any remaining WM capacity used to store the products of reading comprehension, e.g., facts, pronoun referents, and propositions. The two functions, the processing of the written information and the storage of intermediate products, were thought to compete for the limited resources available to WM.

They suggested that individual differences in reading comprehension could be due to variability between readers in the efficiency of their processing skills. Presumably good and poor readers have equivalent overall working memory capacities. Good readers are assumed to have efficient reading skills which demand relatively little from the gross WM resources leaving more of the WM capacity for the storage of products of the reading task. Consequently, good readers have more WM capacity available for the storage of products simply because they have more efficient reading skills. Since poor readers are assumed to have inefficient and thus capacity demanding reading skills, they have less residual WM capacity for storing the products of reading.

Accordingly, individual measures reflecting WM capacity should be tied to a specific processing task, in this case reading. To restate this idea, good readers would have more available WM capacity than poor readers while reading because of their more efficient reading skills. But this greater WM capacity would be specific to reading tasks. That is, a good reader could have LESS WM available when performing a non-reading task than a poor reader who is skilled at the non-reading task.

Daneman & Carpenter (1980) developed a measure of WM that measured capacity while the subject was performing a nominal reading task. This task, which they called the reading span task, was presumed to require both processing and storage in WM. Subjects read groups of sentences aloud while simultaneously trying to remember the last word of each sentence. The number of sentences in a a group gradually increased. After each group, the subject recalled the "endwords" in any order. WM capacity was defined as the largest group of endwords the subjects recalled correctly. The task is, in effect, a word span task since the only measurement made is the number of words recalled correctly. It was hypothesized that the effect of the secondary task (reading the sentences aloud) would differ for good and poor readers. The number of words recalled, against the background of processing sentences, was considered a measure of the residual storage capacity of WM. Daneman & Carpenter (1980) found high correlations between reading span and three measures of reading comprehension; (1) answering fact questions ($r = .72$), (2) pronoun reference questions (r=.90), and (3) the Verbal Scholastic Aptitude Test (Exp. 1, r=.59, Exp. 2, r=.49). In addition, as they had hypothesized, a simple word span test given to the same subjects did not significantly corr late with any of the three comprehension measures. Daneman & Carpenter (1983) replicated the significant correlations between reading span and Verbal SAT (.46 and .58) in two later studies. A similar correlation (r=.53) between WM span and the Nelson-Denny, a standardized test of reading comprehension, was found by Masson & Miller (1983). Recently, Daneman & Green (1986) found that reading span correlated with learning the meaning of novel vocabulary words in a context with sufficient cues for inferring meaning (r(28)=.69). Furthermore, the correlation was still highly significant when vocabulary knowledge effects were statistically removed (r=.53).

From these findings, Daneman and colleagues argued that the reading span is an index of the WM capacity that is NOT allocated to processing (i.e., reading and comprehending) the individual sentences. Because good reading comprehenders have better or more efficient reading strategies than poor reading comprehenders, more capacity would remain for storage of to-be-remembered information. Thus, they argue that any measure reflecting the capacity of a WM that is important in reading comprehension must require the use of reading strategies. The WM span measure is dependent on the type of background task used while measuring the span, and that background task must include reading if the span measure is to predict individual variation in reading comprehension.

Another possible explanation of the Daneman & Carpenter findings, however, may be that people are good readers because they have a large WM capacity independent of the task being performed. Good readers may be remembering more words against the background of processing sentences in the reading span task because they have larger WM capacities than poor readers, NOT because good readers have more efficient reading skills than poor readers. A greater WM capacity would be independent of the type of background task used while measuring span. That is, a good reader may have more WM capacity available for processing and storage than a poor reader whether performing a reading or a non-reading task. According to this alternative theory, a measure of WM should successfully transcend task dependence in its prediction of higher level cognitive functioning. That is, the memory span task could be embedded in a concurrent processing task that is unrelated to any particular skills measure and still predict success in the higher level task.

A pilot study performed in our lab supports the notion that WM storage capacity is independent of the nature of the task being performed (Turner & Engle, 1986). Span measures embedded in a processing task other than reading correlated with reading comprehension just as well as did the reading span. Three span measures were used in the Turner & Engle study. One was a replication of Daneman & Carpenter's (1980) reading span (referred to here as the sentence-word span test, SW) wherein subjects read a series of unrelated sentences and recalled the last word in each sentence. In a second span measure, a to-be-remembered digit followed each sentence in the series, (sentence-digit span test, SD) and the digits were to be recalled following the end of the series. In the third span task subjects performed simple arithmetic operations (e.g., (3 x 4) + 11 = . . .) followed by a to-be-remembered word, (operation-word span test, OW). Memory span was defined as the maximum number of items (digits/words) recalled. The main purpose of this pilot study was testing whether the relationship between these span measures and reading comprehension is dependent or independent of specific processing strategies required by the secondary task. That is, does the secondary task need to involve reading, as Daneman & Carpenter (1980, 1983) suggested, to predict reading comprehension? Daneman & Carpenter's hypothesis would predict that the sentence-word (SW) and the sentence-digit (SD) span tasks would reflect differences in residual WM capacity because of individual differences in reading skills and strategies, not because of differences in WM capacity independent of task proficiency.

The alternative explanation, that the reading span measures abiding individual differences in WM capacity independent of the skills required for the processing component of the task, suggests that the complex memory span index could be embedded in any task that requires heavy processing beyond the span task, and still reflect individual differences in WM capacity that are important in higher level functioning. Using a concurrent processing task involving strings of arithmetic operations followed by a to-be-remembered word (OW), allowed the span task to be embedded in a processing task that requires a different set of strategies than reading comprehension. Nevertheless, the word span measured in the operation-word task predicted reading comprehension just as well as it did in the sentence-word (i.e. reading span) or sentence-digit task.

One implication of these results is that the background task in a WM measure does not need to be reading related to lead to a correlation between the span measure and a test of reading comprehension. An individual may be a better reading comprehender because of a larger WM capacity not specific to more efficient reading skills.

However, there is a possible confounding in the Turner & Engle study. Reading ability and mathematical ability tend to be highly correlated. Thus, good readers also tend to be good in math. Daneman & Carpenter's theory could still be correct by the following analysis. Good readers could have a large sentence-word span because of their efficient reading skills and, independently, they could have a large operation-word span because of their efficient mathematical skills. The dissociation of the nature of the background task and the nature of the task being predicted, in this case, reading comprehension, would only be meaningful when the two types of skills are not themselves correlated.

The purpose of the first experiment was to further address the question of whether WM storage capacity is task dependent. More importantly, the study addressed whether the correlations observed by Turner & Engle are simply due to good readers also having good quantitative skills and using both kinds of skills efficiently, or, whether they generally have larger WM capacities.

One way to approach this problem is to group subjects by their verbal and quantitative abilities and determine whether the operation-word span still correlates with reading comprehension even when the verbal and quantitative scores differ widely. Subjects were divided into four categories based on their Verbal and Quantitative SAT scores as follows: (1) HH, those high in both verbal and math skills, (2) HL, those with high verbal and low math skills, (3) LH, those with low verbal and high math skills, and (4) LL, those low in both verbal and math skills. A comparison of the correlations found between the different WM spans and reading comprehension for the different groups, particularly the HL and LH groups, should help tease apart the shared variance in the correlations due to the correlation of reading and mathematical abilities, from that due to WM capacity.

A second way of approaching the problem of whether these correlations are task specific or task independent is to statistically remove the effects of quantitative skills (as reflected by Quantitative SAT scores) from the correlation between each of the WM span measures with the reading comprehension measures. If the correlation between the operation-word span and reading comprehension is simply due to good readers also having good and efficient quantitative skills, then the correlation between operation-word span and comprehension should disappear when the quantitative skills are factored out. On the other hand, if the correlation between the WM span measure and reading comprehension is independent of the particular skills involved by the background task, the partial correlation coefficient between operation-word span and reading comprehension should still be significant.

METHOD

The Turner & Engle experiment suggested that one of the reasons people are good readers may be that they have a large, general WM capacity that is independent of specific task strategies. The purpose of this experiment was to further test this idea while eliminating alternative hypotheses. It would help to understand the procedure of theses experiments if we consider the complex tasks like Daneman & Carpenter's reading span as really two tasks. The primary task is remembering items, either words or digits. But this memory task is performed against the background of a secondary or processing task like reading sentences. Daneman & Carpenter argue that the significant correlation between the number of items recalled in the span task will only correlate with a reading comprehension task when the secondary or background task is also reading related.

Since different skills are clearly required for solving arithmetic problems than when reading sentences, the secondary or background task of the current study was either solving arithmetic operations or reading unrelated sentences. In addition, since the recall of digits likely requires different memory strategies than those used when remembering words, the to-be-remembered items in the memory component of the task were digits or words.

Therefore, the first experiment included a replication of our pilot study, and in addition, asked whether any similar correlations found between the different complex WM spans and comprehension were simply due to good readers also having good quantitative skills and using both kinds of skills efficiently, or, whether they just generally have larger WM capacities independent of these skills.

Subjects

The study tested 243 University of South Carolina psychology students who participated to satisfy a course requirement or receive extra credit. All subjects were tested in groups of at least two and no more than five and completed the tasks in two one-hour sessions.

Design

All subjects completed seven tasks: four complex WM span tasks, two simple span tasks, and the Nelson-Denny Standardized Reading Comprehension test. In addition, written permission was obtained from each subject to obtain their Verbal and Quantitative Scholastic Achievement Test (SAT) scores from the University.

The four complex memory span measures were: (1) one in which subjects read a series of sentences, verified whether each sentence made sense, and then recalled the last word in each sentence (sentence-word span task, SW), (2) one in which subjects also read a series of sentences and verified whether each sentence made sense and recalled a to-be-remembered digit which followed each sentence, (sentence-digit span task, SD), (3) one in which the subjects verified the answers to strings of arithmetic operations and recalled a to-be-remembered word which followed each operation (operation-word span task, OW), and, (4) one in which the subjects verified the answers to strings of arithmetic operations and recalled the digit answer that was printed to the right of the equal sign regardless of whether or not it was the correct answer (operation-digit span task, OD). There were 42 stimuli in each of the four complex span tasks (SW, SD, OW, OD), presented in 12 trials. The number of stimuli in each trial was gradually increased from two to five, with three trials at each level.

The two simple span measures were: (1) a digit span task in which subjects were asked to remember increasingly larger sets of randomly generated digits in the correct serial order, and (2) a word span task, in which the sets of items to-be-remembered were words rather than digits. There were 132 digits presented in 27 trials, and 81 words presented in 18 trials. The number of stimuli (words/digits) in each trial was gradually increased from two to nine in the digit task, and from two to seven in the word task with three trials at each level in both simple span tasks.

The Nelson-Denny Reading Comprehension Test

Form F of the Nelson-Denny standardized Reading Comprehension test was administered. A total of eight passages, each approximately 250 to 650 words in length, and 36 questions were contained in the test pamphlet. Subjects were instructed to silently read each passage and then to answer the multiple choice questions that followed each passage on the numbered answer sheets provided. They were allowed to look back at the material they had read, but were advised to leave difficult questions until they had finished answering the remaining questions. Subjects were given twenty minutes to complete the test. Reading rate was determined for each subject by having the experimenter call "Mark" when 30 seconds had elapsed. At that time subjects recorded the number of words read, which was printed to the right of the line on which each was reading. Accuracy scores were based on the number of correct answers out of the total 36 questions.

Memory Span Tasks

Word stimuli: The words to be recalled in the sentence-word, operation-word and simple-word tasks were selected from a total pool of 243 words (see Appendix A) and then divided into three sets (i.e., Set A, Set B, and Set C) so that each subject recalled different words in each of the three span tasks. The to-be-remembered word was the last word in the sentence in the SW task, or the word that followed the operation string in the OW task, or was the stimulus item in the simple word task. All three sets of words were one syllable concrete nouns, selected from the most common four to six letter words published in the Francis & Kucera (1982) frequency norms.

Sentence stimuli: Sentences for the sentence-word and sentence-digit tasks were generated to make sense with the designated last-word. Each sentence was from 11 to 16 words long, was nominally unrelated to the others, and was either "correct" or "incorrect". "Correct" sentences made semantic and syntactic sense, e.g., 'The grades for our finals will be posted outside the classroom door.'. "Incorrect" sentences were made non-sense by reversing the order of the last four to six pre-terminal words, e.g., 'The grades for our finals will classroom the outside posted be door.'. The number of "correct" and "incorrect" sentences in a trial (i.e., set size) gradually increased from 2 to 5, with three trials at each set size. Approximately half of the sentences in each set size were randomly selected to be "correct", and half incorrect, with the constraint that 21 sentences in each task were "correct" and 21 were "incorrect".

Digit stimuli: The to-be-remembered digits in the sentence-digit, operation-digit and simple-digit tasks were randomly sampled with replacement from the integers, 1 - 9. A single to-be-remembered digit followed each sentence in the sentence-digit task, or, was the stated answer in the operation-digit task, or was the stimulus item in the simple digit task.

Operation stimuli: A total of 84 operation strings served as stimuli for the processing component of the operation-word and operation-digit complex span tasks. Each string consisted of two arithmetic operations and a stated final answer, e.g., [(9 / 3) - 2 = 1]. The first operation was a simple multiplication or division problem in parentheses, such as (3 x 4) or (8 / 2). The inter-item product or quotient was not stated, and was to be solved prior to a second operation, i.e., the simple addition or subtraction of a single-digit integer. Approximately half of the operation strings in each trial listed an answer after the equal sign that was correct, and half listed an answer that was incorrect by at least 4, e.g.,[(9 / 3) - 2 = 6]. The number of operation strings in a trial (i.e., set size) gradually increased from 2 to 5, with three trials at each set size. As with the sentence stimuli, approximately half of the strings in each set size were answered correctly, and half incorrectly.

General Procedure: In all of the memory span tasks, small groups of subjects read aloud series of stimuli that they saw projected from transparencies onto a large screen while simultaneously hearing pre-recorded items on cassette tape. The length of each recording determined the presentation rate for each stimulus item. The average playback time for the four different stimulus items in the complex tasks was 423 msec for the SW task, 532 for the SD task, 396 for the OW task, and 541 for the OD task. The experimenter kept all items covered on the transparency, except the item being presented, by gradually moving a blank sheet of paper with a 6"x 0.5" cut out, creating a "window" through which the stimulus item was projected onto the center of the screen. Thus, all subjects heard the stimulus items while seeing them on the screen, and were required to read the items aloud, along with the voice on the tape. Subjects performed in groups of from two to five subjects, sitting at individual desks, positioned so that the experimenter could see the mouths of all subjects. Then, if necessary subjects were reminded to read aloud, and in addition, the experimenter mouthed all items, which served as a constant reminder for subjects to continue reading aloud.

The first trial was initiated by the experimenter after reading specific instructions at the beginning of each task. In the sentence-word (SW) task the experimenter allowed the first sentence to be projected through the "window" on the screen, and played the recording of the same sentence. Subjects immediately began reading the sentence aloud paced by the recording. As soon as each subject decided whether a sentence made sense, he/she placed a check mark in the appropriate numbered blank on Side 1 of his/her answer sheet. Immediately after the first sentence was presented, the next sentence appeared on the screen and subjects read it aloud, again paced by the recording of the sentence. After a series of two sentences were read, the subjects saw a red line and heard a "recall" cue at which time they turned their answer sheets to Side 2 and wrote down the last word of each of the sentences they had read. After a set of three trials at set size 2, the number of sentences was increased to three.

And, after each additional set of three trials, the number of sentences continued to be incremented by one until subjects were reading five sentences prior to the "recall" cue. Likewise, in the sentence-digit (SD) task, wherein each sentence was followed by a digit, subjects saw, heard and read each series of sentences and digits aloud, along with the voice on the tape. Again, subjects verified whether the sentence made sense, and when they heard the "recall" cue they serially recalled the to-be-remembered digits. In the operation-word (OW) task subjects also saw, heard and read aloud series of operation strings and to-be-remembered words, again paced by the recorded stimulus presentation. Subjects immediately verified whether each stated answer following the equal sign was correct or incorrect, and finally when the "recall" cue was heard they wrote the to-be-remembered words in any order. In the operation-digit (OD) task subjects saw, heard and read aloud the series of operation strings, paced by the recording, and as soon as possible verified whether the stated answer was correct. When the recall cue was heard they recalled the answer stated after the equal sign in the correct serial order, whether or not the stated answer was correct. In the OW and OD tasks all intermediate computations were required to be silent and no aids (i.e., pencil or paper) were allowed.

In the two simple span tasks subjects saw, heard and read aloud words or digits, sequentially presented at a rate of one per second. The number of words presented prior to the recall cue was gradually increased from 2 to 7, and the number of digits was increased from 2 to 9. When the recall cue was heard, subjects wrote their answers on answer sheets numbered from 1 to 81 for word recall, and from 1 to 132 for digit recall. Digits were recalled in the correct serial order and words in any order.

RESULTS EXPERIMENT 1

The data from this study consisted of 22 scores for each of the 243 subjects, they were derived from the four complex memory tasks, the two simple memory tasks, the Nelson-Denny Reading Comprehension task, and the Verbal, Quantitative and Total SAT tests. The data from the four complex span tasks consisted of one measure of the processing component and two of the storage component of the task. The processing component measure was the number of correct verifications for sentences or operations in each task. The span data for a subject were not included in the analysis if the verification measure was below 80%. The data for two subjects were excluded from each of the complex spans. One recall measure was a Set-Size Memory Span. This was the maximum size of the set of to-be-remembered items (words or digits) in which the subject perfectly correctly recalled the memory items two out of the three times. A second recall measure was a Total Memory Span. This was the total number of correctly recalled words/digits from all trials. Within each trial words could be correctly recalled in any order, while digits had to be recalled in the correct serial order to be included in the total memory span. These two measures were also derived for each of the two simple span tasks.

The scores derived for each subject from the Nelson-Denny Reading Comprehension test consisted of (1) the number of correct answers out of all 36 possible questions, (2) the percentage of correct answers out of the total questions completed, and (3) the number of words read per minute. The second measure was used as an attempt to express comprehension independent of reading speed. However, since this measure did not correlate with the span or any other comprehension measures used in the current study and is not the standardized measure for the Nelson-Denny, it will not be discussed further.

The two span measures led, with few exceptions, to the same conclusions so only the results of the total memory span will be reported. This measure of WM has been used by other researchers (e.g., Baddeley, Logie, Nimmo-Smith & Brereton, 1985 and Masson & Miller, 1983) and will allow better comparison across studies. The highest possible score for this measure was 42 for the complex spans, 81 for the simple word span and 132 for the simple digit span. Table 1 reports the descriptive statistics for the WM measures and comprehension measures.

Table 1

DESCRIPTIVE STATISTICS FOR WM TOTAL SCORE

Tasks

	Sentence Word	Sentence Digit	Operation Word	Operation Digit	Simple Word	Simple Digit
Mean	29.1	32.5	32.6	28.4	63.0	95.8
SD	5.1	5.8	4.4	6.2	10.6	19.2
Minimum	17	18	19	15	23	24
Maximum	41	41	41	40	84	130

DESCRIPTIVE STATISTICS FOR COMPREHENSION AND SAT SCORES

	Nelson/Denny *Number* Correct	Verbal *Reading* Rate	Quantitative SAT Score	SAT Score
Mean	26	269.8	441.5	456.9
SD	5.03	86.3	88.47	91.1
Minimum	8	36	240	270
Maximum	35	756	680	720

Reading Comprehension: Table 2 shows the Pearson Product Moment correlation coefficients calculated between the total memory span and the comprehension measures. The major goal of Experiment 1 was to determine whether the relationship between the WM span scores and reading comprehension varies as the nature of the WM background task was varied in the complex tasks. Table 2 shows the correlational coefficients central to this question. Clearly all four complex span measures correlated significantly with the Nelson-Denny even though the magnitude of that correlation was nearly twice as high for those tasks involving the recall of words as for those involving the recall of digits. The Nelson-Denny correlated significantly with Sentence-word (SW) span, $r(241) = .37$, $p < .0001$, with Operation-word (OW) span, $r(241) = .40$, $p < .0001$, with Sentence-digit (SD) span, $r(241) = .20$, $p < .002$, and with Operation-digit (OD) span, $r(241) = 24$, $p < .002$. These correlations suggest that comprehension may be related to WM span, whether measured against a background task involving reading or performing arithmetic problems. Subjects with higher Sentence-Word spans were better reading comprehenders than subjects with lower Sentence-Word spans. More importantly, subjects with higher Operation-Word spans were also better at reading comprehension than those subjects with lower OW spans. The correlations of Nelson-Denny with Sentence-Word and Operation-Word spans were not statistically different, $t(241) = 0.32$, $p > .50$.

Table 2

CORRELATIONS BETWEEN TOTAL SPANS AND COMPREHENSION MEASURE

Span Task	Nelson-Denny Accuracy Scores	Verbal SAT Scores	Quantitative SAT Scores
Sentence-Word	.37**	.28**	.26**
Operation-Word	.40**	.34**	.33**
Sentence-Digit	.20*	.08	.24*
Operation-Digit	.24*	.11	.24*
Word	.07	.08	.09
Digit	.10	.12	.12

* $r(241)$, $p<.002$

** $r(241)$, $p<.0005$

N=243

The magnitude of the correlation between the two complex digit spans and comprehension was relatively smaller than it was between the two complex word spans and comprehension. The correlations between the Sentence-Digit span and Nelson-Denny, $r(241) = .20$, $p < .0023$, and between the Operation-Digit span and Nelson-Denny, $r(241) = .24$, $p < .0002$, were not significantly different from each other, $t(241) = 0.57$, $p > .50$. Although all four complex spans correlated significantly with reading comprehension, those requiring word recall predicted reading comprehension significantly better ($r = .37$ and $.40$) than those tasks requiring the recall of digits ($r = .20$ and $.24$). Significant differences at the .05 level were found between all combinations of the SW and OW pair of spans and the SD and OD pair of spans; SW/ND ($r = .37$) and SD/ND ($r = .20$), $t(241) = 2.90$, SW/ND ($r = .37$) and OD/ND ($r = .24$), $t(241) = 1.99$, OW/ND ($r = .40$) and SD/ND ($r = .20$), $t(241) = 3.23$, and OW/ND ($r = .40$) and OD/ND ($r = .24$), $t(241) = 2.13$.

Verbal SAT: Differences in the predictability of reading comprehension by the two types of complex spans (digit and word) were also found in the correlations between the complex spans and VSAT scores. Table 2 shows that both complex word spans correlated significantly with VSAT, SW/VSAT, $r(208) = .28$, $p < .0001$, and OW/VSAT, $r(208) = .34$, $p < .0001$. On the other hand, neither of the correlations involving complex digit spans and VSAT were significant, SD/VSAT, $r(208) = .08$, $p > .24$, and OD/VSAT, $r(208) = .11$, $p > .10$. Thus, there was no relationship found between VSAT and complex digit spans, but complex word spans did predict VSAT.

Quantitative SAT: There was a relationship, however, between all four complex spans (SW, OW, SD, and OD) and QSAT. Table 2 shows significant correlations between SW and QSAT, $r(208) = .26$, $p < .0002$, OW and QSAT, $r(208) = .33$, $p < .0001$, SD and QSAT, $r(208) = .24$, $p < .0005$, and OD and QSAT, $r(208) = .25$, $p < .0003$. These significant and nearly equivalent correlations between the four complex spans and QSAT might be expected, considering that the different components of the QSAT would invoke both verbal and quantitative skills, i.e., names of the digits, etc., and that VSAT correlated with QSAT, $r(208) = .54$, $p < .0001$.

Simple Spans: As expected, the simple word and digit spans did not predict reading comprehension. The correlation between the simple word span and Nelson-Denny was, $r(241) = .15$, $p > .097$, and between simple digit span and Nelson-Denny was, $r(241) = .16$, $p > .075$. In addition, Table 2 shows there were no significant correlations between VSAT or QSAT with word span, $r(208) = .08$ and .09 respectively, or with digit span, $r(208) = .12$ and .12 respectively. Therefore, individual differences in the number of words or digits recalled did not predict individual differences in comprehension as measured by Nelson-Denny, VSAT or QSAT scores.

Reading Rate: There was no relationship between reading rate and WM capacity. The only measures that approached significance in correlation with reading rate were the SW and OW complex spans, $r(241) = .17$ and .19 respectively, $p < .06$. No other span and reading rate correlations approached significance.

Span Inter-correlations: The four complex spans tended to correlate with one another, ranging between .35 and .54 (see Table 3). Not surprisingly, the highest span correlations were between the two

Table 3

CORRELATIONS AMONG COMPLEX AND SIMPLE SPANS

Tasks

	Sentence Word	Sentence Digit	Operation Word	Operation Digit	Simple Word	Simple Digit
Sentence Word	–					
Sentence Digit	.50**	–				
Operation Word	.53**	.43**	–			
Operation Digit .38**	.58**	.38**	–			
Word	.10	.03	.08	.15	–	
Digit	.24*	.27*	.35**	.18*	.46**	–

* r(241), p<.01

** r(241), p<.0001

complex spans requiring word recall (SW and OW), r(241) = .53, p < .0001, and, between the two complex spans requiring digit recall, (SD and OD), r(241) = .58, p < .0001. The remaining correlations were also fairly large among the complex spans, SW/SD, r(241) = .50, SW/OD, r(241) = .38, OW/SD, r(241) = .43, and OW/OD, r(241) = .38, at the p < .0001 level. In addition, Table 3 shows a large correlation between the simple word and digit spans r(241) = .46, p < .0001. Thus, the simple spans were highly correlated, and the complex spans highly correlated, however, the correlations across the different types of spans (complex and simple) were much lower, ranging between -.03 and .35. To the extent that the correlations among the complex spans were similar, these memory measures may be tapping the same underlying process. And, of course, the correlation between the two simple spans suggested these two measures may be, at least partially, determined by the same underlying process.

This was, in fact, supported by a principle components analysis of the span scores which accounted for 59% of the variability in the set of span measures. The variable loadings (with eigenvalues greater than 1) of the span measures on each of the components are shown in Table 4. The first component, which accounted for 37.3% of the variance, carries variance associated mainly with the four complex span measures,

Table 4

PRINCIPLE-COMPONENT ANALYSIS: VARIABLE LOADINGS AND COMMUNALITIES

	Component		
Variable	1	2	Communality
Sentence Word	.74	-.18	.58
Sentence Digit	.77	-.18	.63
Operation Word	.74	.03	.54
Operation Digit	.71	-.28	.58
Word	.09	.81	.66
Digit	.51	.68	.74

showing the following high variable loadings: SW = .74, SD = .77, OW = .74 and OD = .71. The simple digit span also loaded to a lesser degree on the first component, .51. The second component, which accounted for 19.7% of the variance, consisted of the two simple span measures showing variable loadings of Word = .81 and Digit = .69.

Comprehension and SAT Inter-correlations: As expected, the correlations among the comprehension and psychometric measures were all large and significant: (1) Nelson-Denny (ND) and VSAT, $r(208) = .66$, $p < .0001$, (2) ND and QSAT, $r(208) = .41$, $p < .0001$, and (3) VSAT and QSAT, $r(208) = .54$, $p < .0001$.

Summary of Correlational Analyses: The correlational analyses showed that comprehension had a relationship with the four complex spans, and that the simple spans did not predict comprehension. In addition, the complex word spans more adequately predicted comprehension than did the complex digit spans. Span inter-correlations also were higher between the two simple spans, and the complex word spans, than those correlations between simple and complex spans. In addition, the relationship between the complex digit spans (SD/OD), and between the complex word spans (SW/OW), was greater than between complex word and digit spans (SW/SD, SW/OD, or OW/SD, OW/OD). These findings were based on tests of differences in the magnitude of correlations.

However, as pointed out by Marascuilo & Levin (1981), what determines the strength of a relationship, is the closeness of points to a regression line. Thus, any two sets of correlations may be found not significantly different (i.e., the two different sets of points, or ellipses, may be equally close to their respective regression lines), but the corresponding slopes of the two regression lines may not be similar. A linear prediction of reading comprehension may be quite different when using one WM complex span, than when using another, even though the strengths of the two linear predictions (i.e., the two correlations between spans and comprehension) are similar.

The regression of Sentence Word with Nelson-Denny and of Operation Word with Nelson-Denny were central to the present study. An analysis was performed to test the hypothesis that the slopes of these two regression lines were parallel. This hypothesis could not be rejected, $F(1,234) = 0.07$, $p > .25$, $MSE = 40.46$. In other words, the two complex spans requiring the recall of words were similar in their ability to predict reading comprehension as measured by the Nelson-Denny.

Linear Relationship between Span and Comprehension: The principal-components analysis confirmed that the relationships between reading comprehension and complex spans reflecting WM capacity should be additionally explored in a regression analyses. One way to approach the problem of whether one, two or all of the four complex spans are needed to predict comprehension, or, whether the spans similarly account for the variation in reading comprehension, is through a stepwise multiple regression analysis using the forward selection technique (Marascuilo & Levin, 1981).

This procedure allowed a different complex span measure to serve as the first, most powerful predictor of comprehension in each of four models. The relative contributions of each remaining complex span in predicting comprehension in each of the different models were then compared. Table 5 shows the results of the stepwise regression analysis.

TABLE 5

STEPWISE REGRESSION WITH THREE COMPREHENSION MEASURES AS CRITERION

Order of Entry	Predictor Variable	Proportion of Variance Accounted for	Model R	F	p
Nelson-Denny					
1	Operation Word	.1603	.1603	45.25	.0001
2	Sentence Word	.0326	.1929	9.53	.0023
1	Sentence Word	.1342	.1342	36.74	.0001
2	Operation Word	.0587	.1929	17.17	.0001
VSAT					
1	Operation Word	.1146	.1146	26.66	.0001
2	Sentence Word	.0154	.1300	3.63	.058
3	Sentence Digit	.0162	.1462	3.87	.050
QSAT					
1	Operation Word	.1101	.1101	25.48	.0001
2	Sentence Digit	.0135	.1236	3.16	.0771

Table 5 indicates the increment in the ND comprehension variability that was accounted for by each span measure when entered into two different models. In the two models, entering SW and OW first, the change in the predicted variance of ND comprehension due to the addition of a third most powerful predictor span was never significant, indicating that SW and OW spans were sufficient predictors of comprehension. Entering OW span first accounted for 16.03% of the variance, while SW added 3.26%. Entering SW span first in this fashion accounted for 13.42% of the variance, while OW added 5.87%.

Table 5 also shows the stepwise regression analysis with VSAT as the criterion variable. When the four spans were allowed to compete in the stepwise procedure OW span was selected first, accounting for

11.46% of the variance in VSAT, with SW span next, accounting for an additional 1.54%, and SD span last, accounting for an additional 1.62% of the final model, $F(3,204) = 11.65$, $p < .0001$. The OW span was the only variable that accounted for a significant proportion of variance in QSAT, 11.01%, $F(1,208) = 25.48$, $p < .0001$.

The question addressed by this study was whether WM capacity is task dependent. Does a WM measure need to be "reading" related to generate a significant correlation between the span measure and a test of reading comprehension? Finding similar correlations between the four complex spans and the ND comprehension measure implies WM capacity transcends task dependency. These results clearly show that the processing component of the WM span does NOT need to be "reading" related to produce a correlation between the span measures and reading comprehension. This suggests individuals may be good or poor reading comprehenders because of a large or small WM capacity, not because of more or less efficient reading skills. However, there is another possible explanation addressed by this study, it may be that these similar correlations are simply due to good readers also having good quantitative skills in which case the significant correlation of Operation Word with the measures of reading comprehension would be accidental in nature. Thus, the correlation between SW and comprehension could occur because, as Daneman & Carpenter argue, the SW task causes reading skills to be invoked and the residual WM capacity is reflected by the number of words recalled. The OW task would lead to arithmetic skills being invoked and, since verbal and quantitative skills tend to be correlated (.54 between VSAT & QSAT in the current sample), the WM capacity reflected by the OW task would tend to be similar to the WM capacity reflected by the SW task. However, WM capacity would still be task dependent. This problem was approached in two statistical analyses of the data, a group analysis and a partial correlation analysis.

Group Analysis: The group analysis consisted of categorizing the 243 subjects into four groups based on their VSAT and QSAT scores: (1) HH, those subjects achieving high scores on both VSAT and QSAT, (2) HL, those achieving high scores on VSAT and low scores on QSAT, (3) LH, those achieving low scores on VSAT and high scores on QSAT, and (4) LL, those achieving low scores on VSAT and QSAT. High VSAT and QSAT scores were defined as those scores that were one half of a standard deviation above the mean of each sample distribution. Low VSAT and QSAT scores were defined as those scores that were one a half standard deviation below the mean of each sample distribution. The sample distribution means and SD's for the VSAT scores were, x = 442, SD = 89, and for the QSAT scores were x = 457, SD = 90.

Thus, the 30 subjects in Group HH had VSAT scores **above** 487, and QSAT scores **above** 502, with the added constraint that the difference between their VSAT and QSAT scores be less than one SD, or less than 90. The 25 subjects in Group HL had VSAT scores **above** 487, and QSAT scores **below** 412, with the added constraint that the difference between their VSAT and QSAT scores be greater than one SD, or greater than 90. The reverse criteria applied to the 44 subjects in Group LH, who had VSAT scores below 397, QSAT scores above 502, and the difference between their QSAT and VSAT scores was greater than 90. Lastly, the 37 subjects in Group LL had VSAT scores below 397, QSAT scores below 412, and the difference between their VSAT and QSAT scores was less than 90.

Table 6 shows the sample statistics (span, comprehension and SAT) for each of the four groups (HH, HL, LH, & LL).

Table 6

DESCRIPTIVE GROUP STATISTICS

Group	SW	OW	SD	OD	Word	Digit	ND	VSAT	QSAT
	Complex Spans		Simple Spans		Comprehension Tests				
High-High									
Mean	30.4	34.8	34.0	28.6	62.2	94.9	30.6	551	567
SD	3.2	3.4	4.7	5.6	10.7	21.1	2.2	43.8	45.4
Min	22	28	21	17	35	41	26	490	510
Max	36	41	41	38	77	123	34	640	680
High-Low									
Mean	29.7	33.3	31.1	29.1	62.8	94.9	28.8	530	399
SD	4.64	5.2	5.84	5.5	9.4	17.6	4.3	82.5	68.0
Min	22	23	21	19	38	66	20	486	280
Max	38	39	41	41	80	120	35	680	413
Low-High									
Mean	29.1	33.0	33.3	30.0	65.8	104.6	25.2	401	532
SD	4.6	5.2	5.8	5.5	9.4	17.6	4.3	65.9	73.6
Min	19	24	19	18	36	62	14	240	485
Max	41	41	41	41	80	130	31	414	720
Low-Low									
Mean	27.8	30.8	31.5	26.6	63.0	96.5	22.5	338	358
SD	5.7	3.6	7.2	6.3	8.2	15.8	4.6	39.7	37.1
Min	18	19	18	18	36	62	12	240	270
Max	39	39	41	41	76	129	31	390	410

As would be expected, the range of each of span and comprehension measures for each of the four groups was more limited than for the entire sample.

In addition, the range of VSAT and QSAT scores was more limited in the HH and LL groups, than in the HL and LH groups. This is important because a greater range of SAT scores defining any specific group would increase the liklihood of a "true" relationship being observed between span and comprehension scores. Thus, it is more likely that correlations calculated between span and comprehension in the HL and LH groups would reflect a "true" association between WM capacity and comprehension, than those from the HH and LL groups. It is also important to note that the means for the reading comprehension scores were higher for the two groups with high VSAT scores (Group HH, $x=30.6$ and Group HL, $x=28.8$), than for the two groups with low VSAT scores (Group LH, $x=25.2$ and Group LL, $x=22.5$). Thus, the Nelson-Denny comprehension scores supported the validity of the categorization procedure.

The group analysis was performed to answer the question of whether the similar correlations between complex spans and comprehension are simply due to good readers also having good quantitative skills, or whether they just generally have larger WM capacities independent of these skills. Table 7 shows the correlations that are important in

Table 7

GROUP CORRELATIONS BETWEEN SPANS AND COMPREHENSION MEASURES

Tasks

	Sentence Word	Operation Word	Sentence Digit	Operation Digit	Simple Word	Simple Digit
High-High						
ND	.41*	-.03	-.04	.22	.05	-.06
VSAT	.16	.00	-.05	.16	.01	.01
QSAT	.32	.23	.23	.43*	-.29	0.06
High-Low						
ND	.59**	.47*	.23	.33	-.10	.32
VSAT	.64***	.30	.04	.17	-.15	.32
QSAT	.65***	.46*	.10	.26	.08	.40*
Low-High						
ND	.23	.38*	.37*	.32*	.07	.29
VSAT	.17	.31*	.15	.08	.07	.14
QSAT	.14	.34*	.22	.12	.05	.13
Low-Low						
ND	.37*	.18	.01	.24	.03	-.03
VSAT	.42*	.44**	-.19	.21	-.19	-.14
QSAT	.41*	.24	.14	.34*	-.29	.10

* $p<.05$ N=30 for HH N=43 for LH

** $p<.005$ N=25 for HL N=36 for LL

*** $p<.0005$

answering this question, those between **OW span and comprehension.** As mentioned above, the most informative correlations would likely be for Groups HL and LH. Hence, the fact that the correlations between OW and ND from Group HH, $r(28) = -.04$, $p > .85$, and from Group LL, $r(36) = .18$, $p > .29$ were not significant was not surprising. However, the correlations between Operation Word and Nelson-Denny were significant for the High-Low group, $r(23) = .47$, $p < .02$, and the Low-High group, $r(41) = .38$, $p < .01$. When comparing these latter two correlations ($r = .47$ and $.38$ standardized to $z = .60$ and $.48$) no significant difference was found, $z = .45$, $p > .33$. In other words, the OW/ND correlation (.47) for Group High-Low, consisting of individuals with good verbal and poor quantitative skills, was similar to the OW/ND correlation (.38) for Group Low-High, consisting of those individuals with poor verbal and good quantitative skills. Thus, the OW and Nelson-Denny correlation was NOT affected by the level of arithmetic math skills. The group analysis suggested that similar correlations between complex spans and comprehension may be reflecting differences in WM capacities, independent of specific skills required in the complex span task.

Partial Correlation Analysis: The partial correlation technique was also used to address the question of whether the significant correlation between Operation Word and reading comprehension measures was an artifact of the tendency for verbal and quantitative skills to correlate. In this procedure, the possibility was considered that the correlations between each of the complex spans and ND comprehension were confounded by the covariance between verbal and quantitative abilities. Therefore, the effect of quantitative abilities (QSAT) was partialled out prior to measuring the relationships between the complex spans and comprehension. Table 8 shows significant values for the

Table 8

PARTIAL CORRELATIONS BETWEEN SPAN AND COMPREHENSION MEASURES

	Comprehension	Measures
Span Measures	Nelson-Denny Accuracy Scores	Verbal SAT Scores
Sentence Word	.25***	.17*

Operation Word	.25***	.20**
Sentence Digit	.10	-.07
Operation Digit	.10	.02
Word	.06	.03
Digit	.09	.02

* r(208), p<.01

** r(208), p<.006

*** r(208), p<.0004

following partial correlations: (1) SW and ND, r(205) = .25, (2) OW and ND, r(205) = .25, (3) SW and VSAT, r(205) = .17, and (4) OW and VSAT, r(205) = .20. All other partial correlations were not significant. The important point is that the relationship between operation span (OW) and reading comprehension (ND) is still present when quantitative skills (QSAT) are factored out of the association. However, neither of the complex digit spans (SD and OD) showed a significant relationship with comprehension after the effect of QSAT was removed. The zero order correlations for SD/ND (r=.20) and OD/ND (r=.24) were half the magnitude of the SW/ND (r=.37) and OW/ND (r=.40) zero order correlations.

In summary, the significant partial correlations between the OW and SW spans and comprehension suggests the complex span measures of WM capacity may be independent of the particular skills involved in the processing component of the span task. The correlation between OW span and reading comprehension was not simply due to good readers also having good quantitative skills.

EXPERIMENT 2

The purpose of this experiment was to study the relationship between the complex WM span and comprehension measures while manipulating the difficulty of the processing component of the span tasks. The findings of Experiment 1 showed that the complex spans reflecting WM correlated with reading comprehension. More importantly, the correlations found between complex spans and comprehension did not depend on the use of specific reading skills in the processing component of the span task. The complex memory span predicted reading comprehension even if it was embedded in a concurrent processing task that was unrelated to reading skills. However, there was no relationship found between comprehension and the traditional span measures in Experiment 1. Why was reading comprehension predicted by the complex WM spans, but not by the simple word or digit span measures? One reason for this may be that individual differences in the simple spans are a result of differences in the use of memory strategies such as chunking, rote rehearsal, phonetic recoding and elaboration. It is unlikely that these same strategies, with the possible exception of phonetic recoding, would be very important to reading comprehension. The complex span measures may more closely reflect the number of "items" that can be represented in the working memory without the aid of rehearsal. The complex span measures may correlate with reading comprehension because the processing component of the task (i.e., reading unrelated sentences or solving operation strings) inhibits the use of these memory strategies. This theory suggests that any method of eliminating the use of memory strategies while measuring WM should lead to a more accurate measure of this limited capacity memory.

One way to test whether the processing component of the complex span task is simply preventing memory strategy use, is to vary the difficulty of the processing component. Theoretically, the span measured while processing very simple sentences, e.g., 'See the dog.', should be similar to the simple span measure. The processing of these simple sentences would not be so demanding that subjects could not make some use of strategies. On the other hand, reading sentences, such as, 'She saw the small dog run behind the truck.', would be more demanding and would minimize strategy use Therefore, spans measured against an easy background task would not likely predict reading comprehension as well as spans measured against a more difficult background task. This suggests that as the difficulty of the processing component of the task increases, strategy use would be less likely, and the resulting relationship between span and reading comprehension would increase. However, when trying to remember items while processing very difficult sentences, such as, 'The young lady with the old man saw the small dog run behind the brand new 1987 Ford truck.', the correlation between span and comprehension may break down. If the background becomes too difficult several things could happen that would diminish the relationship between the complex span and reading comprehension. One is that some subjects would just give up on the background task and allocate all their resources to the span component. So, to some extent, for these subjects the task would be more like the simple span.

Another result of an exceedingly difficult background task could be that span performance would be compressed to the extent that the range of scores would be restricted. And, to the extent that the range of the span scores is restricted, the resulting correlation between comprehension and span would be decreased. Therefore, since spans measured against easy background tasks may not inhibit the use of memory strategies, resulting in a span similar to the traditional simple spans, and spans measured against very difficult background tasks would be restricted in range, a comparison of correlations between comprehension and easy, moderate and difficult complex spans might result in an inverted U-shaped function.

It is also important to note that if the correlation between sentence-word spans and reading comprehension are considered to depend on the use of reading skills in the processing component of the span task, as Daneman & Carpenter argued, then, as the background task is made more difficult the resulting SW span and comprehension correlations would behave in the manner described above. The SW span/comprehension correlations should increase in magnitude, up to that point where the difficulty level restricts the range of the SW span measure. However, if, as Daneman & Carpenter argue, the correlation between comprehension and complex span is determined by specific task skills used in the background task, then when the difficulty of a background task consisting of operation strings is manipulated, the resulting pattern of span and reading comprehension correlations should not be the same as the pattern of correlations between SW span and reading comprehension. Solving operation strings clearly does not require reading skills, yet the OW span measured against the background task of verifying operation strings correlated with reading comprehension ($r = .40$ in Experiment 1). The OW span/comprehension correlation suggested specific task skills did not determine the complex span measures reflecting WM capacity. If a comparison of the behavior of OW span/comprehension correlations with SW span/comprehension correlations should result in similar shaped functions, it would appear that a task independent explanation would be further supported.

METHOD

In this experiment the difficulty of the processing component of the complex sentence-word (SW) and operation-word (OW) span tasks was manipulated. Three levels of difficulty were defined for the processing components of the SW and OW tasks. Unrelated sentences used in the verification component of the SW task were simple, moderately difficult, or difficult, and, the operation strings used in the verification component of the OW task were simple, moderate, or difficult. Thus, there were three SW and three OW span tasks. Reading comprehension was tested with Form F of the Nelson-Denny Reading Comprehension Test. The stimuli and administrative procedures for this standardized test were identical to those described in Experiment 1.

Subjects

There were 52 undergraduate students enrolled in psychology courses at the University of South Carolina who were tested individually in two one-hour sessions. Two of the subjects were unable to complete the difficult OW verification task, and therefore, their data were not used. The remaining 50 subjects completed six complex WM span tasks and one comprehension task. The order in which the seven tasks were completed was randomized across individuals and sessions to balance practice and boredom effects.

Memory Span Tasks

The word stimuli used in the memory component of the SW and OW complex tasks were identical to those used in Experiment 1. Again, three sets of words were used so that subjects would not recall the same words in any one session.

Stimuli for SW task: As described in Exp. 1, the unrelated sentences were generated backwards using words from the three sets of stimulus words as the last word in each sentence. Sentence difficulty was varied in two ways: (1) linguistic difficulty, and (2) the length of the sentences. The linguistic difficulty of sentences has been tested by many researchers. For example, Miller & McKean (1964) found the reaction time required to transform kernal active-affirmative sentences was a function of the complexity of the transformation required. These findings suggested that processing is more difficult for passive and negative than active and affirmative sentences. Therefore, the unrelated sentences in this experiment were active-affirmative (simple), passive-affirmative (moderate) or passive-negative (difficult). The length of the sentences was varied by adding one phrase (for the moderate sentences) or two phrases (for the difficult sentences) to the kernal sentences used in the simple sentences.

Sentence stimuli: There were three sets of sentences totalling to 180 (60 in each of three sets) simple sentences, 126 moderately difficult and 126 difficult sentences. The simple sentences were active and affirmative, consisting of four to five words, e.g., 'People gave their time.'. The moderately difficult sentences were passive and affirmative, consisting of eight to eleven words, e.g., 'Money is given by people at Christmas time.'. The difficult sentences were passive and negative, consisting of from ten to fifteen words, and was one phrase longer than the moderately difficult sentences, e.g., 'Money was not given by people in that state at Christmas time.'.

Half of these sentences made sense and half were made non-sense by reversing the order of the middle two words in the simple sentences, e.g., 'People their gave time.', or, the last four or five pre-terminal words in the moderate and difficult sentences, e.g., 'Money is given Christmas at people by time.', or 'Money was not given by people in Christmas at state that time.'.

Manipulation check on sentence difficulty: A study was conducted to test whether response time to verify the three different types of sentences conformed to expectations. All simple, moderate and difficult sentences used in Experiment 2 were presented on a computer monitor, one at a time, to 24 subjects whose task was to verify whether each of the sentences made sense. Subjects were given reaction time and accuracy feedback after each verification trial and were instructed to maintain at 90% accuracy. Each verification error produced a "beep" from the computer. When there were more than four beeps in a row the experimenter repeated the instructions, stressing that the subject take longer to respond in order to be more accurate. Subjects were most accurate when verifying simple sentences ($x = 98.7\%$ correct), somewhat less accurate when verifying moderate sentences ($x = 93.7\%$ correct), and even less accurate when verifying difficult sentences ($x = 86.6\%$ correct). Subjects were fastest at verifying simple sentences ($x = 1955$ msec), slower at verifying moderate sentences ($x = 2943$ msec), and slowest at verifying difficult sentences ($x = 3947$ msec). Thus, the performance function resulting from the pilot study confirmed that the sentences used as stimuli in Experiment 2 are differentially difficult as described.

Stimuli for OW task: The difficulty of the operation strings in the OW task was varied in two ways: (1) the number of operations in the string, and (2) whether the fraction, 0.5, was used in the string. Each of the simple operation strings (60 in each of three sets) consisted of the addition or subtraction of two single-digit integers followed by an equal sign followed by an answer. All of the addends were randomly selected with replacement from the digits 1-9, with the constraints that the same digit not be used for both addends, and, that the correct sum also be a digit between 1 and 9 (e.g., 3 + 1 = 4; 8 - 5 = 3). Half of the stated sums were correct and half were incorrect by 4.

Each of the 42 moderately difficult operations in each set consisted of two operations with a stated final answer, (e.g., (9 / 3) + 4 = 7). The first operation in the string was a multiplication or a division problem in parentheses, such as (6 / 3) or (2 x 3), followed by the addition or subtraction of a number between 1 and 9. The integers were all randomly sampled with replacement from the numbers 1 - 9 with the constraints that the unstated inter-item product, and the stated final answer, be whole numbers between 1 and 9. Half of the stated final answers were correct and half were incorrect by 4.

Similar to the moderate operations, each of the 42 difficult operation strings in each set consisted of a multiplication/division operation, followed by an addition/subtraction of a number, with a stated final answer. However, the digits in the problem were constrained so that the stated answer or the unstated inter-item product included the fraction 0.5, e.g.,[(2.5 x 2) + 1.5 = 6.5], or, [(7 / 2) - 2.5 = 1]. The final answers were also constrained to be between 0.5 and 9.5, with half correct and half incorrect by 1 or 2.

Manipulation check on operation difficulty: Similar to the sentence verification study, an operation verification study was conducted to test whether subjects differentially processed the different types of operation strings. All operation strings were randomly presented on a computer monitor to 24 subjects whose task was to verify whether each of the stated answers were correct. Subjects were given reaction time and accuracy feedback after each trial and were instructed to maintain 90% accuracy, and urged to slow down when making too many errors. Subjects were most accurate when verifying simple operations ($\bar{x}$=92.78% correct), less accurate when verifying moderate operations ($\bar{x}$=86.2% correct), and least accurate when verifying difficult operations ($\bar{x}$=65.97%). Subjects were fastest at verifying simple answers (x= 1977 msec), slower at verifying moderate answers (x = 2563 msec) and slowest at verifying difficult answers (x =3459 msec).

General procedure

The procedure for administering the six complex WM span tasks was similar to that described in Experiment 1, except subjects were tested one at a time with the series of stimuli presented on an IBM microcomputer. The number of sentences (in the three Sentence Word tasks) or operations (in the three Operation Word tasks) within each 3-trial series was gradually increased from two to five for the moderate and difficult stimuli levels, and from two to six for the simple stimuli levels. Specifically, each subject read the visually-presented unrelated sentences (SW tasks) aloud and verified whether the sentence made sense on their answer sheet. Immediately after a subject finished reading a sentence aloud the experimenter pressed the return key causing the next sentence to appear in the series. The elapsed time each subject required to read and verify each sentence was recorded. When a series of sentences was completed (i.e., one trial) a question mark appeared at which time the subject wrote, in any order, the last words of the sentences just read. Similarly, in the three OW tasks, each subject read aloud a series of operations and to-be-remembered words, verifying whether each stated answer was correct. Immediately after a subject finished reading and verifying an operation string the experimenter pressed the return key and the next string appeared. When a series of operation strings was completed the question mark appeared which cued the subject to recall the words in any order. No limit was placed on the time allowed for the computations, and the time taken to read and verify the problem was recorded. Subjects were warned that the number of stimuli in each trial would increase as they progressed through the series of each task.

RESULTS EXPERIMENT 2

The data for this experiment consisted of scores obtained for each subject from 36 measures of WM span, reading comprehension, and the Verbal and Quantitative components of the Scholastic Achievement Test. Table 9 shows the mean, standard

Table 9

DESCRIPTIVE STATISTICS FOR WM TOTAL SCORE

	Sentence			Operation		
	Simple	Moderate	Difficult	Simple	Moderate	Difficult
Mean	42.1	29.9	26.8	46.5	32.3	29.0
% recall	70	71	64	78	77	69
SD	5.40	3.78	5.04	4.80	4.20	5.04
Mean Span	3.2	2.9	2.4	3.8	3.1	2.5
Minimum	30	20	17	31	18	18
Maximum	55	40	37	56	39	39

DESCRIPTIVE STATISTICS FOR COMPREHENSION AND SAT SCORES

	Nelson Number Correct	Denny Reading Rate	Verbal SAT Score	Quantitative SAT Score
Mean	26.7	289.5	443.4	479.6
SD	5.45	85.6	406.0	107.7
Minimum	12	170	250	270
Maximum	35	544	680	690

deviation, minimum and maximum for the span measures and the reading comprehension measures. In each of the six complex tasks there were two measures of the processing component of the task: (1) mean response time to complete reading and verifying the item, and (2) accuracy stated in proportion of correct verifications. Two dependent measures of span were also recorded for each task: (1) an absolute memory span consisting of the total number of correctly recalled words, and (2) the memory span consisting of the highest set size in which memory items were correctly recalled from two of the three memory trials given at each set size. As in Experiment 1, the analyses of total WM spans are reported.

Verification Statistics: The level of accuracy in the verification component of the complex span tasks was generally high with not much difference across the difficulty conditions. The sentence verification accuracy was 95%, 97% and 94% for simple, moderate and difficult sentences respectively. The operation verification accuracy was 98%, 97%, and 87% for simple, moderate and difficult operations respectively. On the other hand, response time did vary considerably across difficulty condition. The mean response time was 2.80, 4.28, and 5.28 seconds per sentence for simple, moderate and difficult sentences. For the operations, response time was 3.45, 6.01, and 11.14 seconds for simple, moderate and difficult operations.

Reading Comprehension: This study was directed at the question of why reading comprehension was predicted by the complex WM spans (consisting of simultaneous processing and span tasks), but not by the simple word or digit spans (consisting of only the span task). The processing component of the complex WM span may simply prevent the use of memory strategies, and thereby generate a "purer" measure of WM capacity which then better predicts comprehension than the simple span. If the use of memory strategies are inhibited by using a verification task while measuring span, then as the difficulty of the sentence (SW) and operation (OW) verification tasks is increased, the correlation between WM span and comprehension ought to increase up to a point. Table 10 shows the Pearson Product Moment

Table 10

CORRELATIONS BETWEEN SPANS AND COMPREHENSION MEASURES

ComprehensionSentence				Operation		
Measures	Simple	Moderate	Difficult	Simple	Moderate	Difficult
ND	.37*	.47**	.26	.28*	.32*	.23
VSAT	.25	.30*	.27	.18	.11	.07
QSAT	.04	.08	.08	.10	.18	.10

* r(49), p<.05 for span and ND correlations

** r(47), p<.001 for span and VSAT correlations

correlations calculated between the simple, moderate and difficult complex span tasks and comprehension measures that are central to this question.

The level of difficulty influenced the correlation between WM span and reading comprehension, whether the span was measured with the SW or the OW complex span task. The moderate SW/ND correlation, $r(48) = .46$, $p < .001$, was larger in magnitude than the simple SW/ND correlation, $r(48) = .36$, $p < .01$. This difference between moderate and simple correlations (.46 and .36) was significant, $t(48) = 1.99$, $p < .05$. In addition, the moderate SW/ND correlation (.46) was larger than the difficult SW/ND correlation, $r(48) = .26$, $p > .07$ ($t(48) = 3.16$, $p < .01$). Similarly, the moderate OW/ND correlation, $r(48) = .32$, $p < .05$, tended to be larger in magnitude than the simple OW/ND correlation, $r(48) = .28$, $p < .03$, ($t(48) = .88$, $p > .10$), and was larger than the difficult OW/ND correlation, $r(48) = .23$, $p > .11$, ($t(48) = 2.06$, $p < .025$).

It is important to note that the correlations between comprehension and SW span and those between comprehension and OW span behaved similarly across levels of difficulty. One would not expect these correlations to vary in the same systematic manner, if the complex spans reflecting WM are considered to depend on the use of reading skills in the processing component of the span task, as Daneman & Carpenter argued. That is, if reading skills determine the SW span, then as the sentence verification task using those skills is made more difficult, the resulting SW span/comprehension correlations should increase in magnitude, as they did, up to that point where the difficulty level very demanding. However, the OW/ND correlations show a similar pattern across levels of verification task difficulty. And, since the operation verification task (i.e., the processing portion of the OW span task) clearly did not require reading skills, a reading skill or task-dependent explanation simply cannot not account for the OW/ND correlations varying in the same manner as the SW/ND correlations.

The pattern of SW and VSAT correlations across different difficulty levels of span was somewhat similar to the inverted U-shaped patterns for the SW/ND and OW/ND correlations. That is, the moderate SW/VSAT correlation, $r(48) = .33$, $p < .05$, tended to be larger in magnitude than the simple SW/VSAT correlation, $r(48) = .29$, $p < .10$, and also tended to be larger than the difficult SW/VSAT correlation, $r(48) = .27$, $p < .10$. However, the differences between moderate and simple SW/VSAT correlations were not significant, $t(48) = 0.69$, $p > .10$, and differences between moderate and difficult SW/VSAT correlations only approached significance, ($t(48) = 1.34$, $p > .07$. There were no other significant correlations between SW and OW with VSAT and QSAT scores. This included the relationship between Operation Word and VSAT which was significant in Experiment 1. We can only point out that all the correlations were somewhat lower in this study and that all those correlations involving QSAT disappeared in this study.

Analysis of data from the verification task: The top part of Table 11 shows the correlations between the **accuracy** of the verification task and Nelson-Denny, VSAT and QSAT as a function of the difficulty of the verification task. The bottom part of Table 11 shows these correlations for the **verification response time**. In general, both sets of correlations suggest a rather small relationship between verification accuracy and/or speed with the global ability measures. There were only 4 significant correlations with accuracy and 7 with response time.

Table 11

CORRELATIONS BETWEEN ACCURACY OF VERIFICATION AND ABILITY MEASURES

	Sentence			Operation		
Test	Simple	Moderate	Difficult	Simple	Moderate	Difficult
ND	.15	.28*	.08	.05	.16	-.03
VSAT	.10	.30*	.12	-.01	.26	.19
QSAT	.08	.24	.10	.06	.32*	.31

CORRELATIONS BETWEEN VERIFICATION RESPONSE TIME AND ABILITY MEASURES

	Sentence			Operation		
Test	Simple	Moderate	Difficult	Simple	Moderate	Difficult
ND	-.21	-.29*	-.21	-.22	-.25	-.17
VSAT	-.30*	-.33*	-.26	-.34	-.43**	-.25
QSAT	-.27	-.18	-.21	-.21	-.37*	-.29

* $r(49)$, $p<.05$

** $r(49)$, $p<.002$

However, 7 of the 11 significant correlations were for the moderate level of difficulty (4 for sentence verification and 3 for operation verification).

It is notable that the same general inverted U-shaped pattern of verification/comprehension correlations occurred across the three levels of difficulty that had occurred with the span/comprehension correlations. This pattern was found for 11 of the 12 sets of correlations.

An analysis was also performed on the relationship between the verification performance and the number of words recalled in the complex span tasks as a function of difficulty. Only 7 of 72 possible correlations were significant and these fit no noticeable pattern.

Span Inter-correlations: If the SW and OW complex spans are reflecting the same underlying WM capacity, then they ought to inter-correlate. Table 12 shows that the SW and OW spans,

Table 12

CORRELATIONS AMONG COMPLEX MEMORY SPANS

	Sentence			Operation		
	Simple	Moderate	Difficult	Simple	Moderate	Difficult
Sentence Word						
Simple	-.-					
Moderate	.52***	-.-				
Difficult	.37**	.27	-.-			
Operation Word						
Simple	.61***	.48**	.42**	-.-		
Moderate	.50***	.46**	.15	.68***	-.-	
Difficult	.55***	.45**	.22	.65	.71***	-.-

* p<.05

** p<.005

*** p<.0001

at least across the simple and moderate levels of difficulty, correlated significantly. Only 3 of the 15 possible correlations were not significant and all involved a difficult task as a member of the pair.

GENERAL DISCUSSION

The primary question motivating this study was whether the correlation observed between the complex memory span and measures of reading comprehension is a result of a rather specific interaction between the individual and the task being performed at the moment, or, the result of a relatively immutable capacity transcending specific task. If the complex span predicts reading comprehension when measured against reading- or arithmetic-related background tasks, then the measure can be considered to be independent of specific task skills. Experiments 1 and 2 clearly demonstrated that good readers remembered more words and digits than poor readers regardless of whether the background task required reading or arithmetic skills. The SW, SD, OW and OD spans in Experiment 1, and the moderate SW and OW spans in Experiment 2 correlated with reading comprehension whether the span measures were embedded in the arithmetic verification task (OW, OD), or the sentence verification task (SW, SD). One possible explanation of these findings is that good readers also have good quantitative skills leading to a spurious correlation between the operations tasks and reading comprehension. The OW measure was central to this question, and Experiment 1 demonstrated that the OW complex spans for individuals having good reading and poor math skills (HL), and for those individuals having poor reading and good math skills (LH) equally predicted reading comprehension. In addition, the relationship between Operation Word with reading comprehension was still present when quantitative skills measured by the QSAT were factored out of the association. Thus, a complex span reflecting WM capacity does not need to be "reading" related to generate a significant correlation with reading comprehension.

On the other hand, Experiment 1 demonstrated the lack of correlation between reading comprehension and the simple digit and word spans. This led to the question of why the complex spans, but not the simple spans, predicted reading comprehension. Experiment 2 considered the possibility that the complex span is a more accurate measure of WM capacity because it prevents the use of memory strategies, such as grouping and rote rehearsal. If so, then varying the difficulty level of the background task, against which the complex span is measured, should affect the complex span/comprehension correlations. That is exactly what was found in Experiment 2. The SW/comprehension and OW/comprehension correlations were a function of the different levels of difficulty. Further, it is important to note that these two sets of correlations exhibited similar inverted U-shaped performance functions. That is, when the difficulty level of the reading-related (SW) or arithmetic-related (OW) background tasks was moderate, the span/comprehension correlations were higher in magnitude than when the background tasks were very simple, or, were very difficult. The similarity of the OW and SW correlational patterns further supports the argument that the complex span is independent of ties between the background task and reading comprehension.

Why then do complex spans, but not simple spans, predict reading comprehension? One possibility is that performing the background task against which the complex span is measured, limits the subjects ability to use memory strategies such as rehearsal or grouping of the to-be-remembered information. Baddeley & Hitch (1974) argue that terms like articulatory loop and visual-spatial scratch pad refer to existing structures which are part of working memory. However, it may be more reasonable to think of these maintenance components as temporarily activated strategies, as pointed out by Reisburg, Rappaport & O'Shaughnessy (1984). Memory strategies, such as articulation, grouping and rehearsing, are probably not existing structural components of a WM system waiting to be invoked, but temporary processes that are activated within WM when they are needed for the ongoing cognitive task. Therefore, it may be that the background task in the complex span prevents the activation of helpful memory strategies, rather than preventing the use of existing memory structures. Whether the strategies are conceptualized as temporary activations, or existing structural components of WM, individuals use these strategies while performing the simple span task. There is no background task present in the simple span task, and therefore, memory strategies are used to circumvent the capacity limitations in WM, i.e., the number of memory representations individuals keep active at any one time.

Cohen & Sandberg's (1977) findings support the notion that any measure reflecting the WM limitation important in higher level cognition, must dissociate the influence of memory strategies from the span measure. They demonstrated a relationship between IQ and the recency portion of a digit-recall function. In one experiment they gave their subjects a probed serial recall task in which subjects were presented with auditory or visual lists of nine digits. They also varied whether the digits were presented at 1 digit/sec or 4 digits/sec. After presentation the subjects received a cue to serially recall the first, middle or last three digits in the list. And, as expected, they found digits were recalled better at the beginning and the end of the lists, regardless of the rate of presentation. They then tested the relationship between IQ and recall from each of the three portions of the list. The recency portion of the recall function was the only one to show reliable correlations with IQ, with low IQ subjects showing a greater deficit than high IQ subjects on items at the end of the lists of digits. Further, the most reliable correlations occurred between IQ and the recency portion of the recall function when the presentation rate was 4 digits/sec. However, when they tested the correlation between IQ and the primacy recall, they found NO systematic relationship. They suggested that items at the beginning of the list are represented by an empty STM, and that capacity was then available for the use of rehearsal strategies. On the other hand, when the items at the end of the list occur, the subject has already allocated existing resources and this prevents the use of grouping, rehearsal and other memory strategies.

In a later paper, Cohen & Sandberg (1980) included the above data in a factor analysis, from which they argued that the most critical factor determining a relationship between their subjects' STM capacity measures and IQ is whether STM is empty or partially filled at the time of encoding the to-be-remembered items. The important point is that the primacy portion of Cohen & Sandberg's probed serial recall tasks, remembering items from the beginning of a list, can be compared to the simple digit and word span tasks used in Experiment 1, wherein capacity was also available to rehearse the items. And, the recency portion of Cohen & Sandberg's tasks can be compared to the complex spans used in Experiments 1 and 2, in which the background task prevents subjects from using memory strategies to increase the number of items recalled. The implication is that the span measured while the subject is performing any rehearsal preventative task would more accurately reflect the capacity of the buffer.

Why, then, does the size of the correlations between the complex spans and comprehension appear to depend on whether the items to-be-remembered are digits or words? Experiment 1 demonstrated the magnitude of the correlations between comprehension and the complex word spans (SW and OW) were nearly twice the magnitude of those between comprehension and the complex digit spans (SD and OD). Since the same background tasks were used (i.e., verification of sentences in SW and SD, and verification of operation strings in OW and OD), differences in the to-be-remembered items appear to be crucial.

One possibility may be that memory strategies are more available or more automated for digits than for words. Intuitively, it would seem that opportunities to apply grouping strategies to a list of unrelated digits, rather than a list of unrelated words, occur more often in the real world. And, to the extent that strategies are repeatedly used for processing specific types of information, the strategies become automatic processes. If so, strategies used for grouping digits may be simply more automatic than those used for grouping words, and consequently, would be more difficult to inhibit by the background task.

In conclusion, the major theoretical point suggested by the data in these experiments is that the complex span may be a more adequate reflection of working memory capacity than the simple span. To the extent that individuals vary in the limitation to processing imposed by WM, the magnitude of the WM limitation (or capacity) is reflected by a complex span, but not by a simple span. The higher correlations found between complex WM spans and reading comprehension indicate that the complex and simple spans are measuring a fundamental resource, but that the simple span is also reflecting additional underlying resources. Since the capacity of WM can be reflected by the complex span that is independent of skill levels used in the background task, further investigations using this task will help provide additional constraints toward a model of working memory capacity.

Experiment 3

In validating the concept of working memory for real world cognitive tasks, the studies demonstrating a link between working memory and reading comprehension have seemingly prevented what Crowder (1982) called "The demise of short term memory". However, at least one important and nagging question remains. Why do the complex spans correlate with reading comprehension and the simple spans do not? All of the studies reported to date have used, for the simple span task, lists of words presented at a moderate rate of one word per second. Further, the subjects are typically given some form of free recall instructions. This, of course, is quite different from the typical span type study conducted since 1959. Most of those studies used letters or digits, frequently presented at rates faster than 1/sec and almost requiring serial recall. Even when words have been used in span studies, the recall requirement has typically been serial in nature.

The following experiments were conducted to try to better understand what variables might differentially influence both the complex and simple word span.

Method

The subjects were 96 undergraduates who participated for pay or course credit for a single one-hour session. The items for the span task were 2 sets of 100 high frequency one-syllable words from Francis & Kucera (1982). The two sets were matched for frequency. Two hundred sentences, 13-16 words long, were composed each ending in one of the 200 words. Word set was balanced across the two span conditions. The stimuli were presented on a computer monitor. Order of presentation within a set of words or sentences was random for each subject. The words for the simple span task were presented at a one word per sencond rate. For the sentence span task, as soon as the subject had finished reading the sentence, the experimenter progressed to the next one. The order of task was counterbalanced over subjects with half the subjects getting the simple span first and half getting the sentence span first. Regardless of the order of span task, all subjects received the Nelson-Denny between the two span tasks.

As with the previous studies, and as with Daneman & Carpenter's studies, the subjects first received several sets of size 2 (in this case 5 sets), followed by 5 sets of size 3, 5 of size 4, 5 of size 5 and 5 of size 6. Subjects wrote their recall on answer sheets with the appropriate number of lines for each set size. Half of the subjects were given instructions to recall the words in any order but to not recall the last one first. These instructions are the same ones in previous studies. The other half the subjects were given serial recall instructions. They were to write the words in the space corresponding to the order in which the word had appeared but the writing could be done in any order except writing the last word down first. Subjects were monitored closely to make sure they followed directions.

Results

The mean Nelson-Denny, Verbal SAT, simple span score and sentence span score for the free recall group and serial recall group are shown in Table 13. Again, the span scores were analyzed in several ways with no discernible difference between them. The scores reported here are the total of all words correct over all sets.

Table 13

DESCRIPTIVE STATISTICS FOR FREE AND SERIAL GROUPS

	Free	Serial
Nelson-Denny	27.5	26.7
VSAT	463.0	455.2
Simple Span	86.5	81.1
Sentence Span	68.3	63.9

The Nelson-Denny and VSAT scores are typical for our subjects and while the scores for the free tended slightly higher than for the serial condition, they were not significantly higher. The span scores, understandably, were higher for the free condition also. The subjects performing free recall averaged almost 5 words more for both the simple and complex word spans.

The critical analysis for this study was whether the manner of recalling the words in the two span tasks affected the correlation with reading comprehension. This analysis is shown in Table 14.

Table 14

CORRELATIONS BETWEEN SPANS AND READING COMPREHENSION MEASURES

Free

	Nelson-Denny	VSAT
Simple Span	.16	.17
Sentence Span	.39**	.45**

Serial

	Nelson-Denny	VSAT
Simple Span	.29*	.27*
Sentence Span	.41**	.47**

Serial

(with free fecall scoring)

	Nelson-Denny	VSAT
Simple Span	.29*	.28*
Sentence Span	.45**	.57**

* p<.05

** p<.005

For those subjects performing the free recall in the span task the correlations with reading comprehension replicated previous studies. There were significant correlations with Nelson-Denny and VSAT for the complex sentence span but not for the simple word span scores.

For those subjects performing a serial recall in the span task, however, the results were quite different. The correlations for the sentence span were slightly higher than for the free recall condition and, of course, significant. The big difference was for the simple span correlations. The simple span scores significantly correlated with reading comprehension if recall was serial in nature.

While this finding does not explain the difference between the correlations for simple and complex tasks, it does say that we have been underestimating the relationship between simple span and reading comprehension by using the free recall procedure. We still don't know why the difference exists between simple and complex spans or why the simple span measured with serial recall predicts reading comprehension better than that measured with free recall.

Experiment 4

The following study was performed to investigate the influence of order of set size on the correlation between span and reading comprehension. All of the previous studies have started with set size 2 or 3 and gradually increased the size of the set. This, of course, confounds set size with practice, boredom and interference. It is possible that there would be an interaction between the ability of subject and these effects.

Method

The subjects were 144 undergraduates (48 in each of 3 conditions) who were paid for their services. The stimuli were the same words and sentences used in the previous experiment. Three tasks were administered to the subjects, simple word span, sentence word span and Nelson-Denny. The order of word set and task was counterbalanced as in the previous experiment. All stimuli were presented on a computer screen in the manner described previously and all words for the span tasks were written on answer sheets. Recall was to be serial with instructions the same as those given given to those in the serial recall condition in the previous experiment.

The critical manipulation in this study was the manner in which the span items were presented to the subjects. A third of the subjects received the span items in **Sequential** fashion. This is the same way span items have been presented in previous studies. The subjects first received 5 trials at set size 2, then 5 at set size 3 and so on through set size 6. These subjects wrote their recall on answer sheets with the proper number of lines for each set size, e.g., 3 lines for set size 3, etc. A third of the subjects received the span items in what we will call **Random** condition. These subjects received the stimuli blocked by word or sentence but the number of items to be recalled after a given trial (either words or words from sentences) was random with no cue before the trial began. These subjects wrote their recall on answer sheets with 6 lines for each set, regardless of the number actually presented. Another third of the subjects received the span items in what we call the **Cued** condition. This condition was identical to the Random condition except that 750 ms before a given trial began, a digit appeared on the screen indicating the number of items in the set for that trial. This digit flashed for another 750 ms after the set had been presented. Recall was again on an answer sheet with 6 lines regardless of the number of items in the set for that trial.

The descriptive statistics for the various conditions are shown in Table 15.

Table 15

DESCRIPTIVE STATISTICS FOR SEQUENTIAL, RANDOM AND CUED CONDITIONS

	Sequential	Random	Cued
Nelson-Denny	27.8	28.0	27.0
VSAT	467.7	468.9	445.6
Simple Span	81.9	82.1	82.2
Sentence Span	59.1	63.7	61.9

The three presentation conditions scored about the same on the Nelson-Denny and 2 of the 3 were about the same on VSAT. The Cued condition was somewhat lower on the VSAT than the other 2 condition. But it is clear that presentation condition had little or no effect on either the total number of words recalled in the simple span condition or in the sentence word span.

The results of the correlational analysis are shown in Table 16.

Table 16

CORRELATIONS FOR SEQUENTIAL, RANDOM AND CUED CONDITIONS

	Nelson-Denny	VSAT
Sequential		
Simple Span	.39	.42
Sentence Span	.38	.38
Random		
Simple Span	.26	.38

Sentence Span	.34	.55
Cued		
Simple Span	.34	.33
Sentence Span	.32	.44

The first thing to be noticed is that there was little difference between how the simple and sentence span scores correlated with the Nelson-Denny. The correlations with VSAT, while somewhat variable, were, at most, slightly higher for the complex span scores than for the simple span scores.

Experiment 5

Some researchers (e.g., Klapp, et al. (1983) and Brainerd and Kingma (1984, 1985)) have concluded that the resources supporting STM storage (reflected by the simple digit span) and the resources supporting a processing WM (reflected by the complex memory span) are independent. It is possible that the STM span and the complex span measures are reflecting two entirely different systems. However, there is also the possibility that they are simply measuring different components of the same system. For example, they could be measuring the different functional components defined by Baddeley and Hitch (1974). They made a distinction among the functions of 3 different components of WM (i.e., a central executive, an articulatory loop and a visuo-spatial scratch pad). The central executive was considered the attentional component which flexibly functioned in the processing of incoming information and/or the storage of items to-be-remembered. Processing verbal information in a serial manner was considered the major function of the articulatory loop (Baddeley, et al., 1975), and therefore Baddeley (1979) argued that the loop may be the temporal WM component that is specifically used for rote rehearsal. Since serial recall is required for the simple digit- and word-span measure and subjects use a rote rehearsal strategy in this task, it is possible that the to-be-remembered digits are in actuality represented in the articulatory loop and not the central executive. Perhaps the simple digit span is reflecting only one component of the WM system, i.e., the articulatory loop. If so, then, a different underlying resource (mechanism) may be reflected by the simple span (i.e., the articulatory loop), than the resource underlying the complex span measure (i.e., the central executive). It is possible that the WM span measures predict reading comprehension so well simply because the concurrent task portion of the span measurement prevents use of the articulatory loop or other rehearsal procedures.

Two variables have been found to affect simple span measures: phonemic similarity (Conrad, 1964) and word length (Baddeley, Thomson, & Buchanan, 1975). For example, Baddeley, Thomson & Buchanan (1975) found that shorter words led to increased spans and longer ones to shorter spans. Further, they found the spoken duration of the words to be the major predictor of individual spans, not the number of syllables The number of words individuals could articulate in 1 1/2 seconds correlated with their span measures, and from these findings, Baddeley, et al, argued that the articulatory loop was time based. Therefore, it is possible that performance in the simple span task is enhanced by the use of the articulatory loop as a grouping strategy for the to-be-remembered span items. If so, then perhaps the background task used in the complex span measure prevents the use of this strategy, and the resulting complex span more accurately reflects the capacity of the central executive. On the other hand, the background task may prevent the central executive from selecting the more beneficial grouping and rehearsal strategies, i.e., the articulatory loop. In either case, whether the background task prevents the central executive from selecting the best strategy, or, prevents the use of the articulatory strategy, the complex span measure could reflect the number of memory representations that can be active at one time, without the help of memory strategies such as the articulatory loop.

The following experiment was designed to test this hypothesis. We simply wanted to determine whether the complex spans described above are affected by the same variables that affect the simple spans. If the complex spans are measuring the same psychological processes or store as is measured by the simple spans then they should be influenced by the same variables and in the same manner. The series consists of systematically varying the length of the words to be recalled (Experiment 5) and the acoustic similarity in simple word span and sentence word span tasks. Experiment 5 has been completed and we are starting on Experiment 6.

Method

Sixty-four undergraduates were tested individually for slightly less than 2 hours and were paid for their services. All were native speakers of English. Each subject was tested on five different tasks: two simple word span and two sentence word span tasks and the Nelson Denny Reading Comprehension test.

Stimuli for the span tasks were chosen from the Francis and Kucera norms with the following constraints: (1) 162 one syllable and 162 3 or 4 syllables words were chosen and (2) for each one syllable word chosen from the norms, a closely ranked (within 5 ranks) 3 or 4 syllable word was chosen.

The word pool was used to generate the sentences for the sentence word task. Each sentence was composed to end in one of the words from the pool and to be 13-16 words in length. The resulting 364 sentences were then divided into two groups (henceforth called sets A and B) each composed of 81 sentences ending in short words and 81 sentences ending in long words. Half of the sentences were randomly chosen to have the preterminal 5 words scrambled. As in the previous experiments the subject was to verify those sentences that made sense. Nonsense sentences were evenly distributed over each set of sentences.

The individual words for the simple word span and the sentences for the sentence span were presented on a computer monitor and were to be read aloud. The word lists were presented at one word per second and recall began when a question mark was presented. The sentences were paced by the experimenter who progressed to the next one immediately after the last word of the sentence was read aloud. The subject verified whether the sentences made sense on an answer sheet placed on a table in front of the monitor. As with the words, a question mark indicated the beginning of recall of the last words from all the sentences in the set. Subjects were instructed to recall the words from both the simple and complex span tasks in any order but to refrain from recalling the last word first.

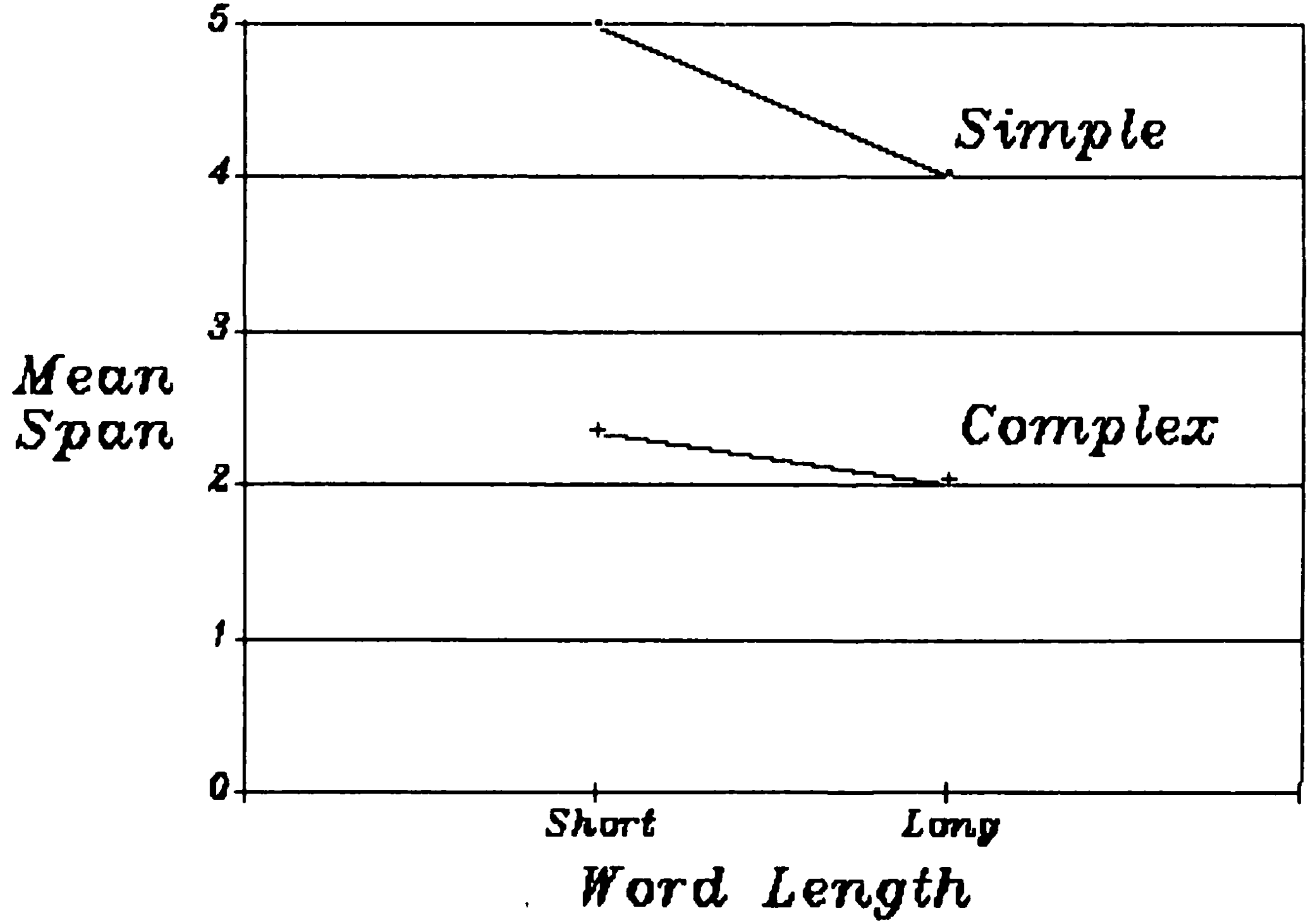
The order in which subjects received (1) simple and complex spans, (2) short and long words and (2) set A and set B words was counterbalanced over subjects. The Nelson-Denny was administered as previously described.

Results

As with the previous studies, the results of the span task were analyzed in several different ways. There were no differences in the conclusions from the different methods. Figure 1 shows the mean average span length for simple and complex tasks as a function of word length.

Insert Figure 1 here

Clearly, there were more words recalled in the simple task than in the complex task and more short words were recalled than long words. There was a standard effect of word length with the simple span task. And, while the effect was smaller, there was also an effect of word length with the complex sentence span task. This is was confirmed by significant main effects of task, $F(1,56) = 940.7$, $p < .0001$, word length, $F(1,56) = 95.3$, $p < .0001$, and the task x word length interaction, $F(1,56) = 27.0$, $p < .0001$. An analysis of just


Simple
Complex
Mean Span
5
4
3
2
1
0
Short
Long
Word Length

the complex sentence span results showed that even though the effect of word length was small it was highly reliable, $F(1,56) = 15.5$, $p < .0002$.

We want to verify these results with another set of stimuli, with the Operation Word task, and with word length as a between subjects variable, but the results suggest that subjects in both tasks may be using articulatory coding as a means of retaining the words. This might mean that the simple and complex tasks do involve at least some of the same processes.

References

Baddeley, A. D. (1986). Working Memory. Oxford Psychology Series No. 11. Oxford: Clarendon Press.

Baddeley, A. D. & Hitch, G. (1974). Working Memory. In G.H. Bower (Ed.), The Psychology of Learning and Motivation, 8, NY: Academic Press.

Baddeley, A. D. & Lieberman, K. (1980). Spatial working memory. In S. Domic (Ed) Attention and performance, Vol 6, 521-530. Hillsdale, N.J.: Erlbaum.

Baddeley, A. D., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. Journal of Memory and Language, 24, 119-131.

Baddeley, A. D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short-term memory. Journal of Verbal Learning and Verbal Behavior, 14, 585-589.

Brainerd, C. J. & Kingma, J. (1984). Do children have to remember to reason? A fuzzy-trace theory of transitivity development. Developmental Review, 4, 311-377.

Brainerd, C. J. & Kingma, J. (1985). On the independence of short-term memory and working memory in cognitive development. Cognitive Psychology, 17, 210-247.

Broadbent, D. E. (1958). Perception and communication. Oxford: Oxford University Press.

Case, R. (1974). Mental strategies, mental capacity, and instruction: a neo-Piagetian investigation. Journal of Experimental Child Psychology, 133, 386-404.

Chase, W. G. & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), The Psychology of Learning and Motivation, Vol. 16. New York: Academic Press.

Cohen, R. L. & Sandberg, T. (1977). Relation between intelligence and short term memory. Cognitive Psychology, 9, 534-554.

Cohen, R. L. & Sandberg, T. (1980). Intelligence and short-term memory: a clandestine relationship. Intelligence, 4, 319-331.

Coltheart, V. (1972). The effects of acoustic and semantic similarity on concept identification. Quarterly Journal of Experimental Psychology, 24, 55-65.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Daneman, M. & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 561-583.

Daneman, M. & Green, I. (1986). Individual differences in comprehending and producing words in context. Journal of Memory & Language, 25, 1-18.

Francis, W. N. & Kucera, H. (1982). Frequency Analysis of English Language. Boston, Mass.: Houghton Mifflin.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. Cognitive Psychology, 10, 302-323.

Howard, D. (1983). Cognitive Psychology: Memory, Language, and Thought. New York: Macmillan.

Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Klapp, S. T., Marshburn, E. A. & Lester, P. T. (1983). Short term memory does not involve the "working memory" of intellectual processing: The demise of a common assumption. Journal of Experimental Psychology: General, 112, 240-264.

Marascuilo, L. A. & Levin, J. R. (1983). Multivariate Statistics in the Social Sciences: A Researcher's Guide. Monterey, CA: Brooks/Cole.

Martin, M. (1978). Memory span as a measure of individual differences in memory capacity. Memory & Cognition, 6, 194-198.

Masson, M. E. J., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. Journal of Educational Psychology, 75, 314-318.

Miller, G. A. & McKean, K. O. (1964). A chronometric study of some relations between sentences. Quarterly Journal of Experimental Psychology, 16, 297-308.

Nelson, M. S., & Denny, E. D. (1973). The Nelson-Denny Reading Test. Boston: Houghton Mifflin.

Newell, A. & Simon, H. A. (1972). Human problem solving. Englewood, Cliffs, N.J.: Prentice-Hall.

Perfetti, C. A. & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.) Cognitive processes in comprehension. New York: John Wiley & Sons.

Reisberg, D., Rappaport, I. & O'Shaughnessy, M. (1984). Limits of working memory: the digit -span. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 203-221.

Salame, J. & Baddeley, A. D. (1982). Disruption of short term memory by unattended speech: implications for the structure of working memory. Journal of Verbal Learning and Verbal Behavior, 21, 150-164.

Schneider, W. & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), The Psychology of Learning and Motivation, Vol. 21, New York: Academic Press.

Shannon, C. E. & Weaver, W. (1949). The mathematical theory of communication. Urbana, IL: University of Illinois Press.

Turner, M. L. & Engle, R. W. (1986). Working memory. Proceedings of the Human Factors Society, 30, 1273-1277.

van Dijk, T. A. & Kintsch, W. (1983). Strategies of Discourse Comprehension. New York: Academic Press.

Waugh, N. C. & Norman, D. A. (1965). Primary memory. Psychological Review, 72, 89-104.

Wechsler, D. (1944). The measurement of adult intelligence. Baltimore: Williams & Watkins.

Publication Information

The report of the first two experiments is ready for submission and will be sent **Journal of Memory and Language**. The citation should be:

Turner, M.L. & Engle, R.W. Working memory capacity: An individual differences approach. **Journal of Memory and Language**.

The report of experiments 3 & 4 will hopefully be ready to be submitted by late spring and will probably be submitted to **Memory and Cognition**.

The data from experiments 1 & 2 were reported at the meeting of Psychonomics Society this past November.

Goals for next year

The goals for next year are: (1) To publish the results of experiments 1 & 2 in one manuscript and, later, to publish the results of experiments 3 & 4 in a second manuscript. (2) To complete the series of studies on whether the complex task is affected by word lenght and acoustic similarity. This will probably involve 2-3 additional experiments. (3) To further attack the question of why the complex span better predicts reading comprehension than the simple word span task. This will involve several experiments varying the nature of the background component in the complex task and several experiments varying the nature of the simple task to determine if changes in the task can make it as predictive as the complex task. (4) To explore to what extent the background task must be like the reading comprehension task. It clearly does not have to be reading related since the Operation Word task involved arithmetic. But the operation component is still probably verbal in nature. Maybe the background component of the complex span task must be verbal in nature. This would speak to the possiblity that we may need to think about separate "capacities" for verbal and spatial tasks. (5) To attack this same question from the spatial side. We will soon begin an experiment in which the complex word span task will have a background task requiring the classification of symbols of a non-verbal nature. This task can be varied in difficulty so that we can determine whether the same relationship holds between the difficulty dimension and the correlation with reading comprehension as was observed in Experiment 2.

END

DATE

FILMED

6-88

DTIC